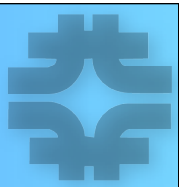


FNAL CMS Software and Computing

Ian Fisk

FNAL DOE Review

May 17, 2006

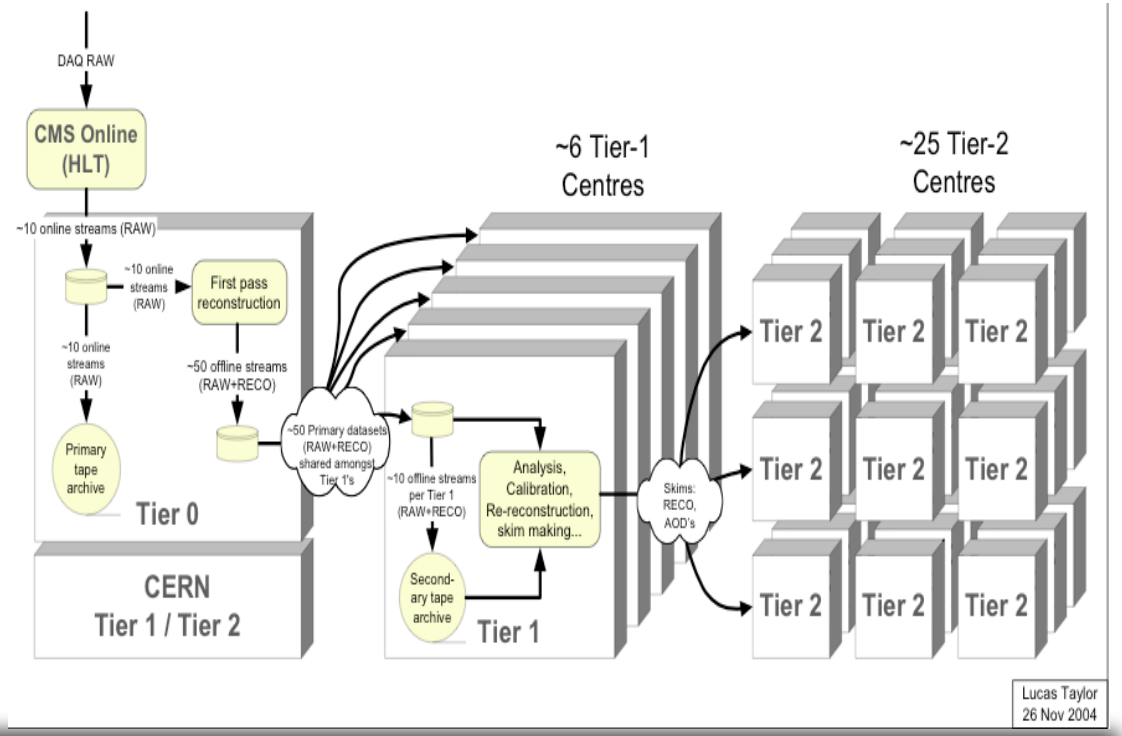
CMS Computing Model

CMS is implementing a distributed computing model where the site activities and functionality is largely predictable

- Activities are driven by data location
- Opportunistic computing is largely restricted to limited activities

High level of distribution and the concentration of resources away from the experiment emphasizes the importance of the computing tiers

- Tier-1s are a natural extension of the on-line
- The need for distributed computing tools drives development in the project



CMS S&C Activities at FNAL

Fermilab is driving several critical elements of CMS preparation

- FNAL is delivering the single largest Tier-1 facility for CMS
  - Its size is specified by the number of authors in the US on CMS
    - Approximately one third of the total collaboration and a center which roughly corresponds to two nominal Tier-1 centers
- FNAL is either responsible for or contributing to a number of the service development activities that are needed to allow the distributed computing facilities to efficiently function as a computing system
  - Data Management Components and Analysis workflow on OSG
  - The service framework for Monte Carlo simulation at Tier-2 centers and eventually re-processing at the Tier-1 centers
  - Integration, development, and deployment efforts within Open Science Grid for CMS
- FNAL is driving the development of the new software framework for CMS
  - CMSSW will be the basis for reconstruction and analysis activities

Responsibilities of the Tier-1

Tier-1 Centers serve as an extension of the experiment on-line computing

- Share of raw data for custodial storage
  - Only active copy of raw data distributed to Tier-1 centers, CERN Tier-0 copy is largely unaccessible due to lack of resources at Tier-0
- Data Reprocessing
  - CMS anticipates 2 reprocessing runs per year

They are entrusted with serving the data entrusted to them

- Selecting and Skimming data for User Analysis and Calibration Tasks
- Data Serving to Tier-2 centers for analysis

Tier-1 centers also support Tier-2 centers

- Some specific operational support responsibilities
  - FNAL supports 7 US Tier-2 centers
- Archival Storage for Simulation and Important Analysis products from Tier-2 centers
  - Tier-2 centers typically do not have tape-based mass storage

The Tier-1 Center at FNAL

FNAL is a dedicated Tier-1 Facility for CMS

- Meeting the obligations of the U.S. to CMS Computing
  - Supporting the local community
- The only Tier-1 center in the Americas

FNAL Tier-1 2008	CPU	4.3MSI2k	1000 dual CPU nodes
	Disk	2PB	200 Servers (1600MB/s IO)
	Network	15Gb/s	CERN to FNAL
	People	30FTE	Includes Developers and Ops

The Tier-1 effort at FNAL is ~10FTE

FNAL is in the middle of the second year of a three year procurement ramp in preparation for the start of the experiment

- First year was relatively smooth from a hardware standpoint with few surprises
  - Several scaling elements will be interesting this year

Facility Services

Grid Interfaces:

- FNAL Spans the boarder between the US and European grid infrastructure (supports both LCG-2.7 and the OSG-0.4 releases)
  - Two doors into the same physical hardware
    - Cluster utilization is roughly half grid submission and half local jobs

Processing:

- All resources were switched to a Condor based system in 2005
  - Cluster is scaling well. Priority scheduling allows reasonable allocation of resources.

Storage:

- dCache/Enstore deployed for Mass storage
  - The dCache system has performed well under heavy load
    - Over 200TB delivered to applications in a single day

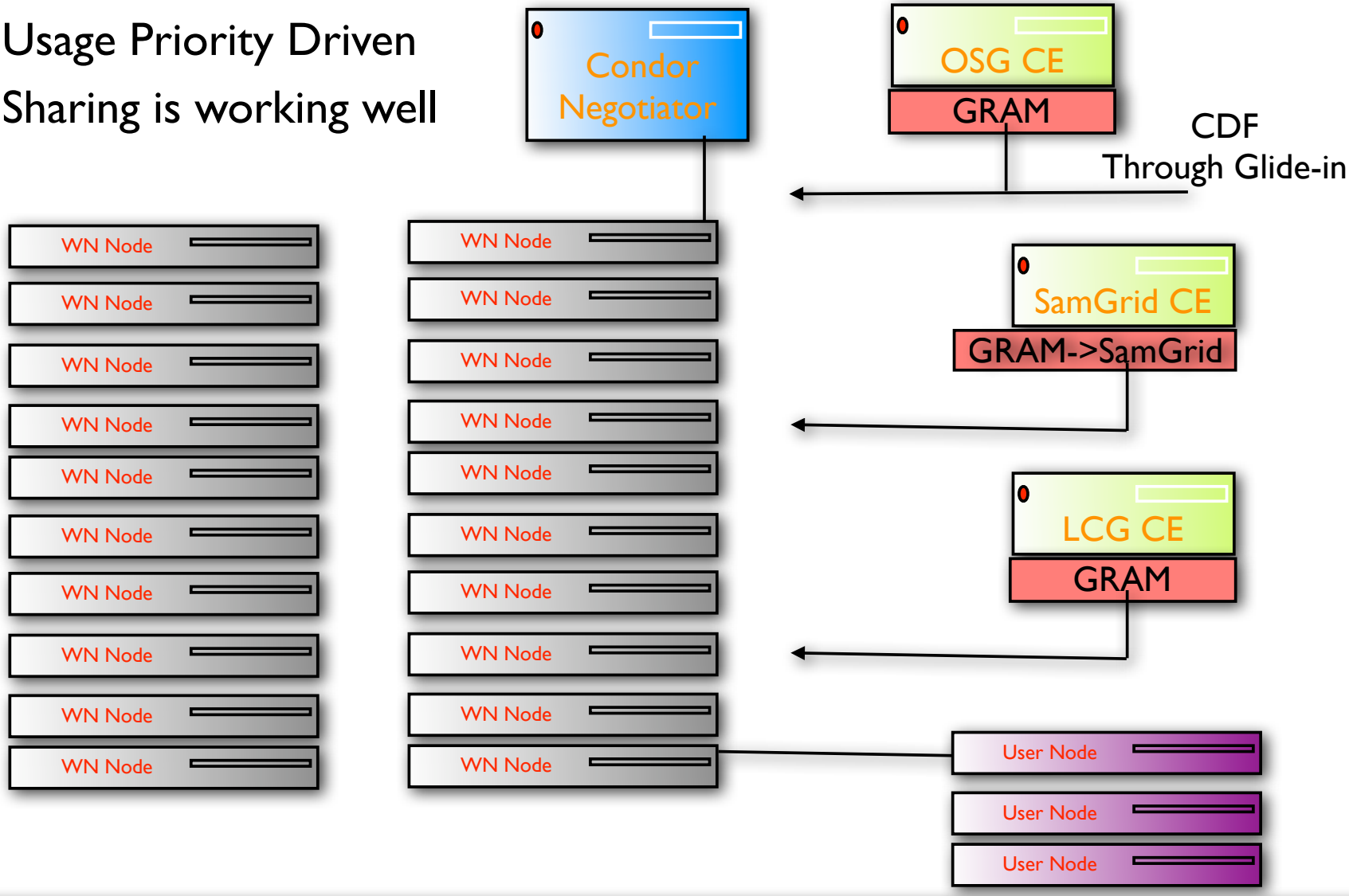
Networking:

- Current we have a 10Gb research link

Grid Deployment

LCG and OSG have individual gatekeepers

- Usage Priority Driven
- Sharing is working well



Processing Resources

FNAL is currently at ~500 dual CPU nodes (1000 CPUs)

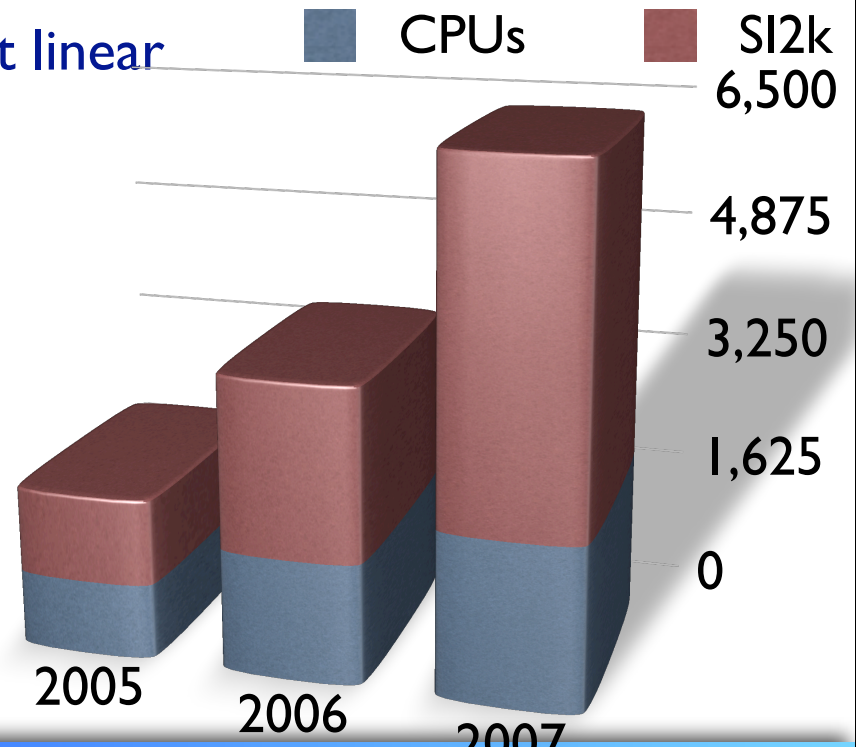
- Slowest are 2.4GHz Xeons and the Fastest are single core Opteron 268s
- Facility is ~1000kSI2k (25% of the expected capacity in 2008)

The operational ramp to the start of the experiment is manageable

- Experience at FNAL configuring and running farms this size

The increase in number of nodes is almost linear

- Performance increase is a fairly conservative improvement estimate
- Dual cores are improving the situation
  - 2006 orders are in



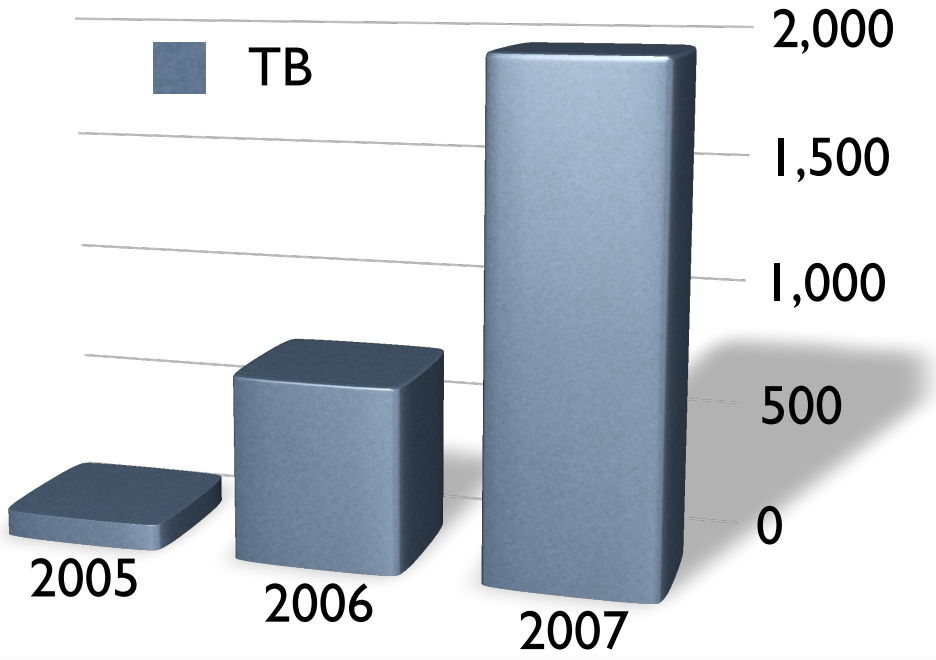
Storage Resources

Currently the FNAL Tier-1 has ~100TB of dCache storage

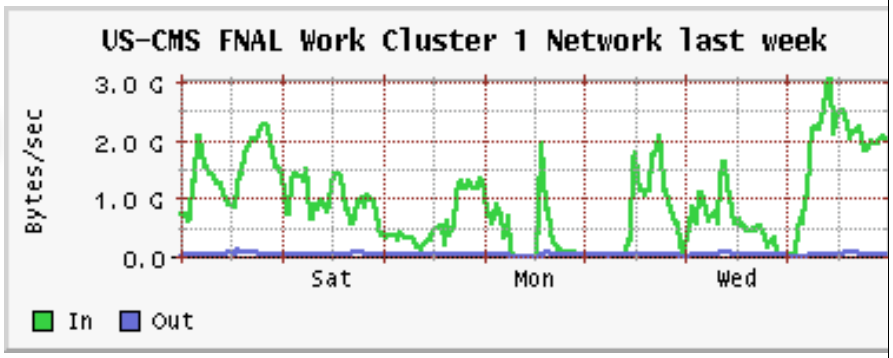
- Roughly 5% of the capacity expected in 2008, but 20% of the number of servers

Very steep operations ramp in disk storage before the experiment start

- Procuring, deploying and commissioning at a large scale
- First half of 2006 requisitions should arrive shortly



Performance numbers for dCache are higher than the schedule



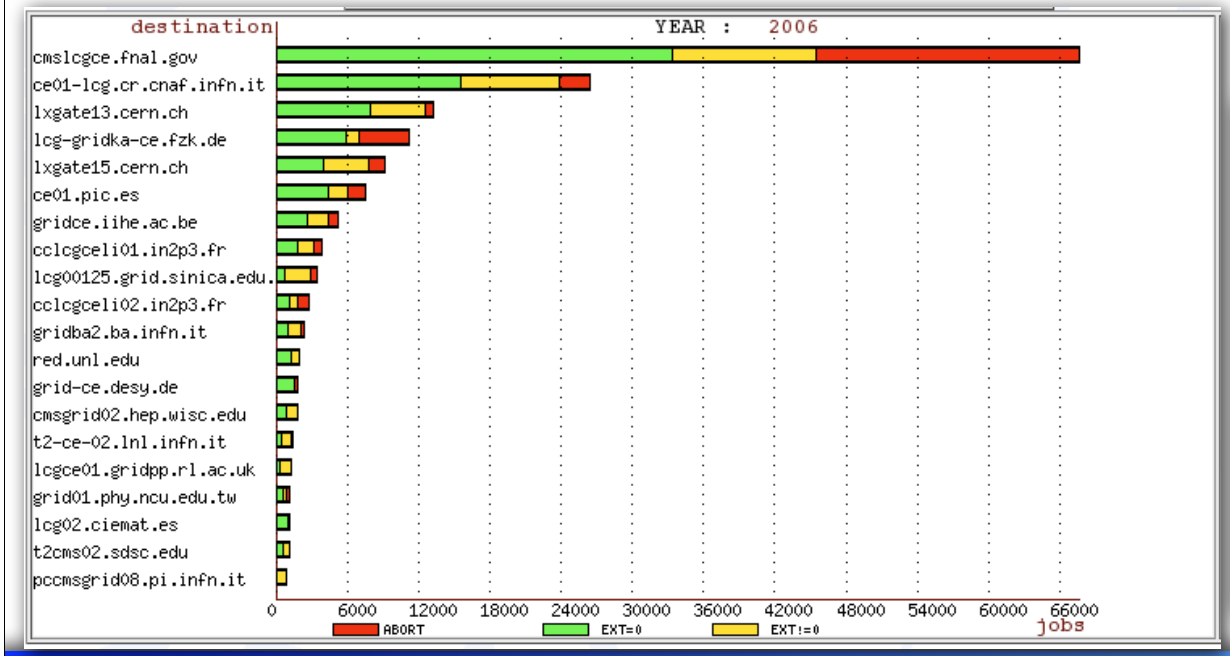
Growing User Load

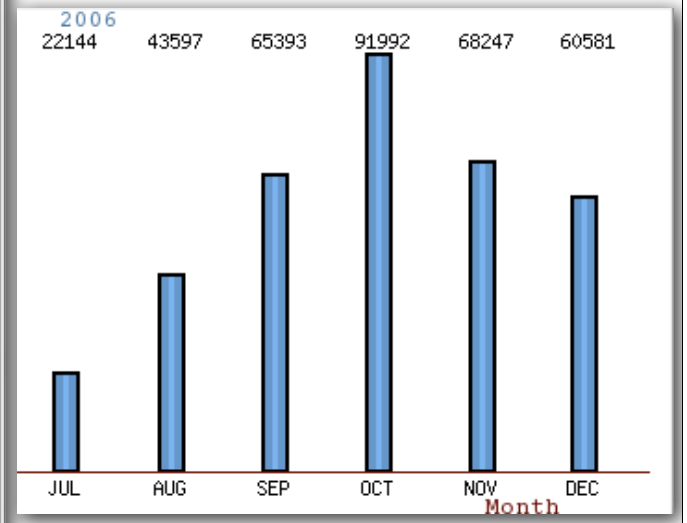
There are over 450 individuals signed up for interactive access to the CMS farm at FNAL

- Somewhat ahead of our projected ramp for users.
- At any given time 10-15% of those are actively running jobs.

The Facility is gaining expertise supporting a growing number of user grid submissions

- Single largest site in CMS for grid analysis submission



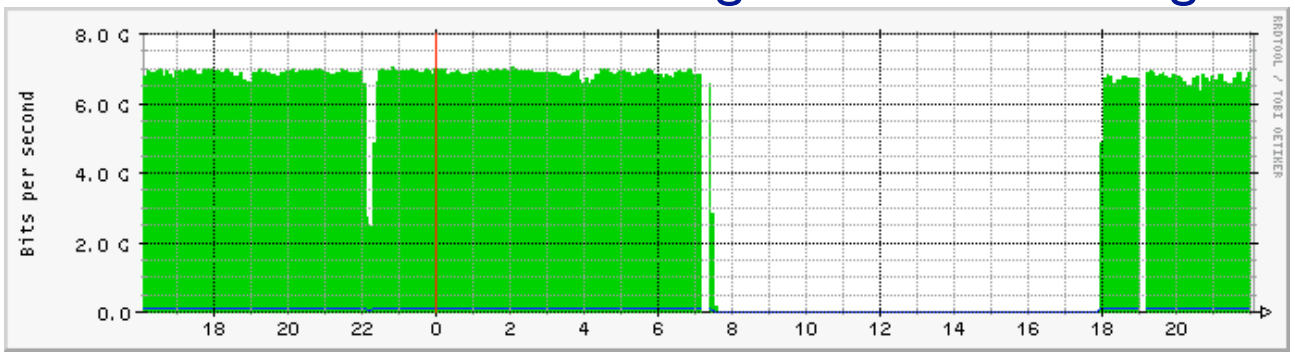


Network Resources

The WAN networking for CMS is provided by a 622Mb/s production link and a 10Gb/s research link (second 10Gb/s link is available to light).

- Most of the CMS data traffic goes over the research link

Traffic between CERN and FNAL during a Service Challenge



The LAN networking is provided by 10Gb/s links between large switches

FNAL network group has provided excellent support

- Working on networking from FNAL to US-Tier-2 centers
- Beginning to measuring network limitations to remote centers
  - Tier-1 and Tier-2

Service Deployment

In order to make the 35 CMS sites currently located on 4 continents function coherently as a computing system new services are needed

- Data Management components
  - Dataset Bookkeeping (meta data), Dataset location, and Data Transfer
    - What is the data I want and where is it?
- Workflow Components
  - Bulk production of simulated events at 25 Tier-2 centers
  - Organized reprocessing of events custodially stored at Tier-1 centers
  - User friendly and supportable methods of supporting analysis jobs to grid enabled remote sites
- Grid service development, integration, and deployment
  - The US relies on the Open Science Grid to connect the Tier-1 and Tier-2 centers
    - Developing interoperability, accounting, security, and troubleshooting

Integration of Data and Work

In CMS the data management functionality and the work load management functionality are tightly integrated

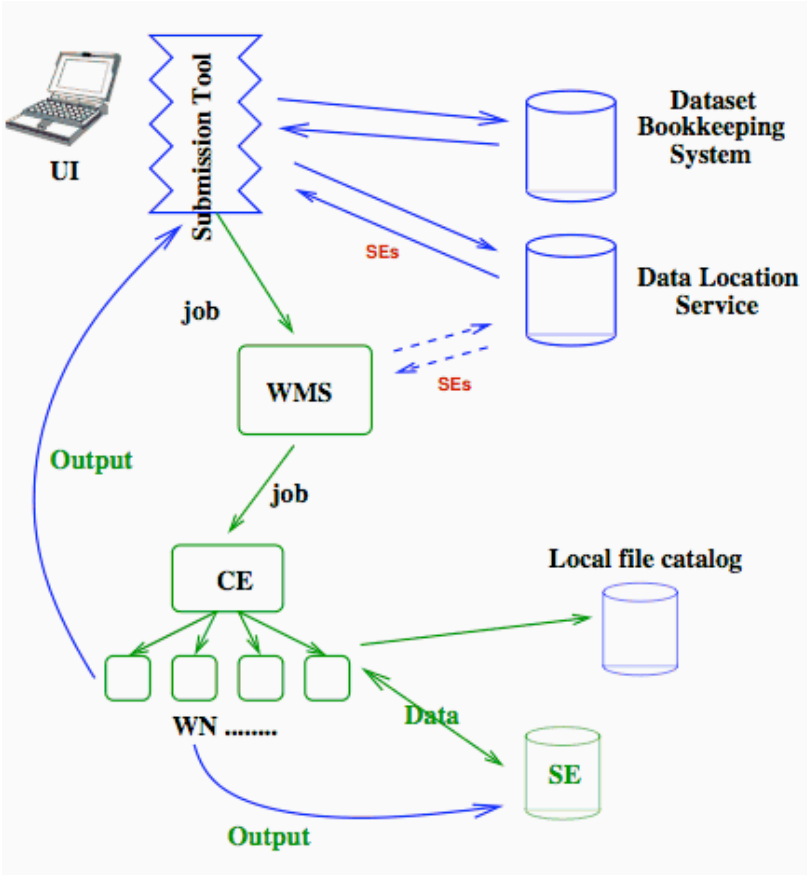
- The data location drives the placement of work

FNAL has contributed to the dataset bookkeeping system (DBS)

- Developing the schema and APIs
- Supporting the current prototype
- Directing project effort at Cornell
  - ~4FTE

FNAL has also driven the analysis workflow deployment on the OSG

- CMS Remote Analysis Builder (CRAB)
  - 0.5 FTE



Data Management Prototype

The 5% data challenge in 2004 (DC40) exposed a number of deficiencies

- A requirements group was formed in the summer of 2004
  - Report recommended changes in data layout and approach to data storage
    - Influences data management and the SW framework for reconstruction and analysis

The new data management prototype has three elements

- Dataset Bookkeeping Service
  - The meta data catalog for the experiment
    - There is currently a prototype implementation. Performs the basic
- Dataset Location Service
  - A european delivered component
- Transfer system
  - CMS (Physics Experiment Data Export) PhEDEx system
  - FNAL supports integration and deployment with US-CMS contributing to development

Environment for Analysis

FNAL has been working on the analysis workflow both in a local and a global context

Locally Goal is to establish a first class environment for analysis in the US

- Serve Datasets
  - FNAL took responsibility to serve all of the JetMet data (about 25%) of the total and had many of the Higgs samples
- Reliable Mass Storage
  - 200TB of tape resident datasets
- Sufficient processing resources and user storage resources

Globally FNAL is working to enable transparent analysis on distributed grid resources

- Enabling US submission to Open Science Grid resources efficiently
- Enabling worldwide submission to OSG transparently

Status of Analysis Activity

At the Tier-1 FNAL is currently serving over 250 simulated data samples

- Ranging from 10k events to 5M events in size

There are over 50 CMS software distributions installed for users

- Simulation, reconstruction, visualization, and framework packages
- Pre-releases of new software framework
- Debugging release of specific tools

There are over 450 individuals signed up for interactive access to the CMS farm at FNAL

- Somewhat ahead of our projected ramp for users.
- At any given time 10-15% of those are actively running jobs.

There is 10TB of user controlled disk space

- Some quota controlled space, some physics group managed space
- Heavily accessed

Remote analysis on OSG

In July of 2005 CMS introduced the CMS Remote Analysis Builder (CRAB)

- A system in which a user could specify the data set desired, the application and input parameters to run, and the the number of events to process per job
- CRAB would handle the data discovery
  - Through a group of catalogs published by the sites
- The job preparation
- Submitting the application

All US Tier-2 sites can be successfully utilized with CRAB

- The official CMS distribution has capabilities for OSG submission
- We have run more than 5000 jobs in a day across the US-CMS Tier-2 sites
- Tier-2s are now visible through the LCG Resources Broker

Production Systems

FNAL has been a driving the developing a production submissions for organized activities like simulation and distributed event reconstruction

- Very large increase in scale and stability
  - 4 sites to more than 20 sites
  - Factor of 10 increase in utilized CPUs for processing

We have leveraged OSG opportunistic computing at universities, national labs, and experiment installations. Made excellent use of large installations like UW

The current system is in deployment

- It has been used to produce a few million events on a combination of local, OSG, and LCG resources
  - The Production Agent module will be deployed for distributed production and the 50M event run for the next data challenge

The New CMS Framework

One of the highest profile S&C contributions at FNAL is the new framework (CMSSW)

- It has been deployed for the Magnet Test and Cosmic Challenge (MTCC)
- It is being deployed for the final CMS data challenge in the fall
- It is the basis of the "Startup Scenarios" for Volume 3 of the Physics Technical Design Report
- As of May 1 CMSSW is the default package for all new CMS analysis

Physics validation will proceed until the end of June

Beginning in July and August large scale distributed analysis at the level of 25M events per month will commence

September and October is the Computing Software and Analysis Challenge (CSA06)

Framework Schedule

There are ~4FTE of effort at FNAL in the new framework and software. With 8FTE in US-CMS as a whole

- The project has an aggressive schedule with major releases every month until the early Fall
  - Currently the full detector simulation is available and many reconstruction elements
  - In June a framework for 10M events for physics validation will be released
  - In July the code needed for the Computing and Analysis challenge (CSA06) will be releases
  - In August full tools for alignment and calibration are expected

Outlook

From the International CMS standpoint the largest profile item has typically been the Tier-1 center

- It is a large fraction of the resources and the Tier-1 has contact with a number of people in the user community. Important resource for CMS
- Progress on the Tier-1 is on schedule to hit the design capacities in time for the start of physics running

US-CMS has always been a large provider of software engineering effort to CMS

- US-CMS was responsible for the visualization program, the production workflow tools, and projects for data persistency and detector geometry
- The concentration of developers at FNAL has resulted in FNAL becoming a nucleus for development activity
  - CMSSW
  - Data Management Elements and Production Tools
  - Software achievements will rival the Tier-1 center for profile